

REMARKS

CONSIDERATION OF REFERENCE

With the Office Action, the Examiner returned a form PTO-1449 for an IDS filed August 3, 2005. On the form PTO-1449, the Examiner that all of the references had been considered except reference AH:

Westendorf, et al., "Automatische Generierung von Ausspracheworterbuchern aus Signaldaten," Proceedings of the 4th Conference on Natural Language Processing-Konvens-98, Computers Linguistics, and Phonetics Between Language and Speech, Peter Lang, Frankfurt Am Main, Germany, 1998, pp. 213-225.

This reference appears in PAIR and is properly cited. Applicants suspect it was not marked as considered because it is in German. However, Applicants note that the second page of the reference includes an English abstract. This abstract represents Applicants' understanding of the reference and Applicants respectfully request that the reference be marked as considered based on the abstract.

CLAIMS 1-8

Claims 1-3 and 8 were rejected under 35 U.S.C. §102(b) as being anticipated by Heckerman et al (U.S. Patent No. 6,263,308, hereinafter Heckerman). Claim 7 was rejected under 35 U.S.C. §103(a) as being unpatentable over Heckerman. Claims 4-6 were indicated as being allowable if rewritten in independent form including all the limitations of the base claim and any intervening claims.

With the present amendment, Applicants have amended claim 1 to include the limitations 2-4. As amended, claim 1 represents claim 4 rewritten in independent form. Based on the indication that such a claim would be allowable, claim 1 and claims 5-8 which depend therefrom are in condition form allowance.

CLAIMS 9-18

Claims 9-15, 17, and 18 were rejected under 35 U.S.C. §102(b) as being anticipated by Heckerman. Claim 16 was rejected under 35 U.S.C. §103(a) as being unpatentable over Heckerman in view of Bartosik et al. (U.S. Patent No. 6,725,194, hereinafter Bartosik).

Independent claim 9 provides a computer-readable storage medium having computer-executable instruction encoded thereon that when executed by a computer cause the computer to perform a series of steps. Those steps include receiving text for a word for which a phonetic pronunciation is to be added to a speech recognition lexicon and receiving a representation of a speech signal produced by a person pronouncing the word. The text of the word is converted into at least one text-based sequence of phonetic units and a speech-based phonetic sequence of phonetic units is generated from the representation of the speech signal. The phonetic units of the at least one text-based phonetic sequence and the speech-based phonetic sequence are placed in a search structure that allows for transitions between phonetic units of the text-based phonetic sequence and phonetic units of the speech-based phonetic description. A phonetic pronunciation is selected from the search structure wherein the selected phonetic pronunciation comprises phonetic units of the speech-based phonetic sequence that differ from phonetic units of the at least one text-based phonetic sequence and phonetic units other than phonetic units of the speech-based phonetic sequence.

Support for the amendments to claim 9 is found in the specification on page 21, line 24 - page 22, line 4, and Figure 6, where it is shown that a selected phonetic pronunciation comprises phonetic units from a speech-based phonetic sequence that differ from phonetic units of a text-based phonetic sequence as well as other phonetic units that are not part of the speech-based phonetic

sequence.

As amended, claim 9 is not shown or suggested in Heckerman or the combination of Heckerman and Bartosik. In particular, neither reference shows or suggests selecting a phonetic pronunciation from a search structure where the phonetic pronunciation comprises phonetic units of a speech-based phonetic sequence that differ from phonetic units of text-based phonetic sequences and phonetic units other than phonetic units of the speech-based phonetic sequence.

In columns 13-15 of Heckerman, Heckerman describes a phonetic transcription system that forms a set of phonetic transcriptions for a text-based word as well as a phonetic recognizer that decodes an audio signal to form a phonetic representation of the word. In column 14, lines 53 - column 15, line 9, Heckerman indicates that a phonetic representation selector compares the phonetic representation generated from the audio signal to the phonetic representation generated from the text. When there is a reasonable match between the speech-based phonetic representation and one of the text-based phonetic representations, the text-based phonetic representation that best matches the speech-based phonetic representation is selected. Thus, when there is a strong enough match, the phonetic representation based on the text alone is selected. In such a case, the phonetic sequence does not include phonetic units that differ from phonetic units of the text-based phonetic sequence as required by amended claim 9.

Column 15, lines 10-19 of Heckerman indicates that when there is a significant difference between the speech-based phonetic representation of a word and the text-based phonetic representations of the word, the speech-based phonetic representation is used instead of any of the text-based phonetic representations. When this occurs, the phonetic pronunciation does not include phonetic units other than phonetic units of the speech-based phonetic

sequence as required by claim 9.

Thus, Heckerman does not provide a situation where a selected phonetic sequence includes both phonetic units from a speech-based phonetic sequence that differ from phonetic units of text-based phonetic sequences and phonetic units other than phonetic units of a speech-based phonetic sequence.

Bartosik also does not provide for the selection of such a phonetic pronunciation. First, Bartosik does not deal with phonetic units but instead deals with entire words. Second, in the graph of Figure 4, Bartosik is not selecting a sequence of words to represent a sentence. Instead, Bartosik is comparing a text-based sentence to a speech-recognized sentence to determine which words were properly recognized by the speech recognition system. This allows Bartosik to identify which audio samples can be used for retraining. Bartosik assumes that the text provided is correct since it is based on a user correction. Thus, under the teachings of Bartosik, the text is always selected as being correct. Applying this teaching to Heckerman would lead those skilled in art to always select one of the text-based pronunciations and to not select the phonetic representation of a word generated by a phonetic recognizer. In any case, Bartosik does not show selecting a phonetic pronunciation that includes phonetic units of a speech-based phonetic sequence that differ from phonetic units of a text-based phonetic sequence as well as phonetic units other than phonetic units of a speech-based phonetic sequence.

As noted in the present specification, it is the ability to integrate phonetic sequences identified for part of a word from a text-based system with phonetic sequences identified for another part of a word for a speech-based system that has not been accomplished before. This allows difficult words such as "voiceXML" to receive a proper phonetic description that cannot be obtained by using either a text-based phonetic description or a speech-based

phonetic description alone. The combination of Heckerman and Bartosik does not show or suggest the problem solved by the invention in claim 9 or the solution suggested in claim 9. As such, claim 9 and claims 10-18, which depend therefrom, are patentable over Heckerman and Bartosik.

CLAIMS 19-27

Claims 19-23 were rejected under 35 U.S.C. §103(a) as being unpatentable over Heckerman in view of Bartosik. Claims 24-27 were indicated as being allowable if rewritten in independent form.

Claim 19 provides a method for adding an acoustic description of a word to a speech recognition lexicon. A text-based phonetic description is generated based on the text of a word and a speech-based phonetic description is generated without reference to the text of the word. The text-based phonetic description and the speech-based phonetic description are aligned in a structure. The structure comprises paths representing phonetic units with at least one path for a phonetic unit from the text-based phonetic description being connected to a path for a phonetic unit from the speech-based phonetic description. A sequence of paths is selected and an acoustic description of a word is generated based on the selected sequence of paths. The acoustic description comprises a phonetic unit found in the speech-based phonetic description but not in the text-based phonetic description and a second phonetic unit found in the text-based phonetic description but not in the speech-based phonetic description.

The amendments to claim 19 find support in the specification on page 21, lines 24 - page 22, line 4 and in Figure 6. In particular, Figure 6 shows a selected sequence of paths that includes a phonetic unit found in a speech-based phonetic description but not in a text-based phonetic description and a second phonetic unit found in the text-based phonetic description

but not in the speech-based phonetic description.

As amended, claim 19 is not shown or suggested in the combination of Heckerman and Bartosik. In particular, neither reference shows or suggests generating an acoustic description of a word that includes a phonetic unit found in a speech-based phonetic description but not in a text-based phonetic description and a second phonetic unit found in the text-based phonetic description but not in the speech-based phonetic description.

As noted above, in Heckerman, either a speech-based phonetic description or a text-based phonetic description is selected. Heckerman does not show mixing a speech-based phonetic description and a text-based phonetic description for a word. Bartosik does not describe phonetic descriptions of words. In addition, in Bartosik, the text provided by the user is assumed to be correct. As such, Bartosik does not teach mixing phonetic units found in a speech-based phonetic description with phonetic units found in a text-based phonetic description to produce an acoustic description of a word. Instead, using the teachings of Bartosik, one skilled in the art would simply use the text-based descriptions since the text is assumed to be correct in Bartosik.

Since neither Heckerman nor Bartosik show or suggest selecting an acoustic description of a word that comprises a phonetic unit found in a speech-based phonetic description but not in a text-based phonetic description and a second phonetic unit found in a text-based phonetic description but not in a speech-based phonetic description, claim 19 and claims 20-27 which depend therefrom are patentable over the combination of Heckerman and Bartosik.

CONCLUSION

In light of the above remarks, claims 1 and 5-27 are in form for allowance. Reconsideration and allowance of the claims is

respectfully requested.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to deposit account No. 23-1123.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By: 

Theodore M. Magee, Reg.No. 39,758
Suite 1600 - International Centre
900 Second Avenue South
Minneapolis, Minnesota 55402-3319
Phone:(612) 334-3222 Fax:(612) 334-3312

TMM:sew